DOCUMENT RESUME

ED 092 175	95	IR 000 742

AUTHOR	Pinsky, Paul D.
TITLE	Achievement Monitoring of Individually Paced Instruction. Final Report.
INSTITUTION	Sequoia Union High School District, Redwood City, Calif.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C.
PUB DATE	Oct 73
CONTRACT	OEC-9-72-0012
NOTE	31p.; Study conducted at San Carlos High School
EDRS PRICE	MF-$0.75 HC-$1.85 PLUS POSTAGE
DESCRIPTORS	Achievement Tests; *Computer Assisted Instruction; Criterion Referenced Tests; Earth Science; Field Studies; *Formative Evaluation; High School Students; *Individualized Instruction; *Secondary Grades; Summative Evaluation; *Test Construction
IDENTIFIERS	*Comprehensive Achievement Monitoring

ABSTRACT

A study was made to monitor achievement of individually paced instruction. The project concentrated on designing testing procedures in group paced instructional programs to provide information to student, teachers, parents and administrators which could be used in both a formative and summative evaluation. The three objectives of the project were: (1) to adapt the Comprehensive Achievement Monitoring (CAM) design for an individually paced program of instruction that contains a series of units through which students progress in sequence; (2) explore the applicability of computer-assisted instruction evaluation technique to criterion referenced testing (CRT) for individually paced instruction; and (3) to field test the adopted CAM design in a high school earth science course. The results showed quite strongly that the students whose learning activities were controlled the most showed the greatest gains in achievement levels. Gains were measured by the CAM tests and by standardized tests given at the beginning and end of the course. The results show this population of students are not able to work independently with CRT data and direct their own study activities. This finding confirms less formal studies completed in previous years. (WCM)

ERIC

ED 092175

TR

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

FINAL REPORT

ACHIEVEMENT MONITORING OF INDIVIDUALLY PACED INSTRUCTION

OEC-9-72-0012

Paul D. Pinsky

October, 1973

TM 003 742

CONTENTS

1. Introduction

The report discusses achievement monitoring of individually paced instruction. During the past decade many educators have advocated changing the educational system to provide instruction that is designed to better serve the individual needs of students (1,2). A common practice used in many schools today is the individual pacing of students through a fixed sequence of content material (3). This technique of instruction, however, presents certain problems concerning the evaluation and control of the learning process. The most common mechanism for controlling the instructional activities of students is through a pretest-posttest design. In this design the student takes a criterion-referenced test (CRT) on the unit of material before instruction has occurred. Based upon the results of the pretest, he directs his study activities. When the student feels that he has learned the material in the unit, he takes a parallel posttest. If he masters the posttest, he goes on to the pretest for the next unit; if he does poorly on the posttest, he continues to review the objectives in the unit.

The pretest-posttest control mechanism, however, does have certain disadvantages. For example, the amount of clerical work necessary to run such a program is substantial. Most programs using the above testing procedures have several forms of each unit posttest, and the instructor must provide the correct form for each student. Further, the pretests and posttests are usually processed by members of the instructional staff. Several projects throughout the country are using computer technology to help alleviate this clerical problem (4,5). The initial findings, however, indicate that even with a computer system, the cost of operating such a testing program is high.

The recently completed Comprehensive Achievement Monitoring (CAM) project explored the use of sampling techniques to help provide the types of information described above (6). The project concentrated on designing testing procedures in group paced instructional programs to provide information to students, teachers, parents and administrators that could be used in both a formative and summative fashion. The CAM system is currently being used in over 100 school districts throughout the country.

The three objectives of the project are to:

a.) Adapt the CAM design for an individually paced program of instuction that contains a series of units through which students progress in sequence. The adaption should provide more information for decision-making about individuals and at the same time retain some of the information about groups of students.

b.) Explore the applicability of computer-assisted instruction (CAI) evaluation technique to criterion referenced testing (CRT) for individually paced instruction.

c.) Field test the adopted CAM design in an earth science course at San Carlos High School, San Carlos, California. A major subobjective is to determine the extent to which students can make their own decisions regarding instructional activities in an individually paced program without the usual pretest-posttest design.

ERIC

The adaption of the CAM design for individually paced instruction has been completed and is presented in Section 2. The evaluation designs are based upon the decisions that are to be supported when the data is generated. Educational decisions have been divided into three categories or levels.

Level I:	Decisions about individual students based on diagnostic test data, on a pretest, posttest, or retention basis.
Level II:	Decisions about instruction on a course, program, class, or curriculum basis.
Level III:	Decisions about schools, districts, regions, and the entire state.

Three evaluation models, the unit CAM, the sliding unit CAM, and the standard CAM are presented. Terminology is defined that enables one to characterize a variety of individually paced evaluation designs in terms of the basic models. The three models are compared relative to the type of information they generate for Level I and Level II decision making.

The application of CAI evaluation techniques to CRT situations is discussed. From a practical viewpoint, the analysis is not very encouraging. The major advantage of the CAI concept is that a history of student performance is stored in the computer. This performance history is used to determine the test items to present to each student. Thus, testing is done dynamically to account for individual student differences and growth over time. The individualized CAM designs are static in that they treat all students who are working on the same unit in the same manner. That is, the CAM technique does not account for differences in student learning patterns within the same unit of material.

There are several problems in applying CAI techniques to CRT. The first is that the computer would have to print a customized test for each student at each test administration. Problems arise in entering the test items into the computer; in printing special symbols, characters, and diagrams on existing computer output devices; and in the cost of generating such tests. Moreover, the cost of data entry into the computer, and the administrative costs of keeping track of the large numbers of documents going back and forth to the computer would probably be prohibitive.

The most promising approach to applying CAI evaluation techniques to the CRT situation might be to do on-line CRT. The cost of computer power and the cost of terminal devices (such as cathode-ray tubes - CRT) drops every year. Therefore, an area for future research would be to devise an on-line CRT system that accounts for individual student differences.

The field testing of the modified CAM design produced some interesting results. Because of small sample sizes and circumstances perhaps unique to the San Carlos study body, the results must be considered tentative and should be subjected to additional studies in different environments. The study contained 3 classes of approximately 28 students each. One class was free to chose their own study materials and to decide when they had learned these materials. Another group was directed in their study efforts, but still had some decisions concerning when they had learned the material. The third group worked closely with the instructional staff concerning what they were to study and when they had learned the material.

The results show quite strongly that the students whose learning activities were controlled the most showed the highest gains in achievement levels. These gains were measured by the CAM tests and by standardized tests given at the beginning and end of the course. The students in the course tend to be underachievers in the science area. Many took the earth science course to fulfill the high school graduation requirement in science. The results show quite clearly that this population of students are not able to work independently with CRT data and direct their own study activities. This result confirms less formal studies that have been completed in the same course in previous years.

2. Evaluation Designs for Individually Paced Instruction

Decision Levels

This part of the report discusses different evaluation designs for the use of criterion-referenced testing (CRT) data. These evaluation designs reflect the variety of education decisions that are made using the results of CRT data. The type of evaluation design used should depend upon the decisions that will be made when the results are generated.

One must recognize that there is a wide variety of decision makers in an educational enterprise. A few examples are teachers as a faculty, teachers as individuals, students, parents, principals, school committees, etc. Each of these people or groups makes decisions about the same educational enterprise. Each, however, makes different kinds of decisions about the enterprise, from different perspectives. Because of this fact, each needs different kinds of data.

For purposes of this report, educational decisions are divided into three categories or levels:

Level I: Decisions about individual students, based on diagnostic test data, on a pretest, posttest, or retention basis.
Level II: Decisions about instruction on a course, program, class, or curriculum basis.
Level III: Decisions about schools, districts, regions, and the entire state.

Each level of decision making needs different kinds of data because each involves different kinds of decisions by different levels of decision makers. This can be better seen by examining some possible decisions on each level.

Examples of Level I Decisions

(1) Have the students in the class mastered the prescribed subject matter?
(2) What objectives of instruction did the students know prior to instruction?
(3) What learning did the students retain after instruction?
(4) What students need additional work on what objectives?
(5) What students do not need to go through a particular learning sequence since they can already perform the skill to be taught?

These Level I decisions need data about how individual students do in relation to specific learning objectives (prespecified criteria).

Examples of Level II Decisions

(1) What objectives should be added to the curriculum? (or deleted? or modified?)
(2) Which objectives are none of the students meeting? Why?
(3) What instructional materials and programs work better in terms of student outcomes at each stage? For what students?
(4) Which instructional modes work in having students achieve which objectives?
(5) Which class(es) are succeeding (or failing) with respect to the objectives in a certain course or curriculum?

These might be some typical Level II decisions. It can be seen immediately that this is a larger level of decision making in that the focus is no longer on the individual student, but rather on groups of students. Each of the above decisions needs group data rather than individual student data. This does not mean that the instruction must be "group-paced" (indeed, the report is focusing on individually-paced instruction), but that the results must be summarized over the individual students to facilitate the Level II decision.

Examples of Level III Decisions

(1) To what degree are the pupils of the state attaining the goals toward which public education is directed?
(2) To what degree are the pupils of each district attaining the goals toward which public education is directed?
(3) Which districts are attaining unusual success and what factors appear to be responsible for that success?
(4) When new educational programs are introduced into the schools, do subsequent changes in pupil accomplishments indicate that the program is accomplishing its purposes?

These kinds of decisions deal with large numbers of individuals, so large that the collection of individual student data could swamp a decision maker. In fact, this has been one of the problems with state-wide evaluation to date. The evaluations have been narrowly conceived, based on misperception of the kind of data needed to make the decisions. There has been confusion between evaluation on the three different levels, or a lack of awareness that the three different levels, demanding three different levels of data, existed.

The report focuses on decision Levels I and II for individually-paced instruction. Evaluation designs for the Level III decisions should make extensive use of sampling techniques (7). This sampling might include sampling of districts, schools, buildings, or students, and extensive sampling of the content domain (item or matrix sampling techniques).

A Hypothetical Curriculum Structure

A carefully designed course structure facilitates the explanation of evaluation designs. The curriculum in the report is idealized, but contains the fundamental ingredients of most curricula that a teacher would devise for an individually-paced program. The course is used to present specific examples of the concepts presented in the report.

The hypothetical course contains eight units, learning activity packages, modules, etc. These units are numbered 11, 12, 13, 14, 21, 22, 23, and 24. The students are expected to learn the units at their own pace in the fixed sequence. The average student will spend approximately 2 weeks on each unit. Six objectives are included in each unit. A four digit identification number for each objective consists of the unit number as its first two digits; e.g. Objectives 1101, 1102, 1103, 1104, 1105, and 1106 are in Unit 11, and Objectives 2301, 2302, 2303, 2304, 2305, and 2306 are in Unit 23. There are a total of 48 objectives in the curriculum because there are eight units with six objectives per unit. There are six test items per objective. The six-digit test item identification number consists of the objective number as its first four digits and a sequential identification number unique to the objective as its last two digits: e.g., Items 210201, 210202, ..., 210206 are related to Objective 2102. There are a total of 288 test items because there are 48 objectives with six items per objective in the curriculum. Note that each item is related to one and only one objective, and that each objective is related to one and only one unit.

Three evaluation models, the unit CAM model, the sliding unit CAM model, and the standard CAM model are presented. In each of these models, a student responds to 10 test forms, one at the completion of each unit, and one at the beginning and end of the program, These ten test administrations for each student are numbered from 1 to 10. Test Administration 1 is the pretest of the entire program, and Test Administration 10 is the posttest. Test Administration 2 occurs as each student completes Unit 1; Test Administration 3 occurs as each student completes Unit 2, etc.

The hypothetical course has an enrollment of 240 students distributed amoung eight classes of 30 students each.

The number of test forms used differs with each model. However, each form in all models contains 24 items. Therefore during the individualized program each model generates exactly the same number of studdent responses to test items. These three models can thus be compared by the quality of information rather than by the quantity they produce. However, the report does not suggest that all test forms should contain 24 items, or that all test forms in an evaluation design should contain the same number of items. The curriculum used in the report is hypothetical and is designed to facilitate the explanation of alternative evaluation designs for individually-paced instruction.

The Unit CAM Model

The unit CAM model consists of a pretest of the entire program during Test Administration 1, and a posttest of the entire program during Test Administration 10. During each of Test Administrations 2-9, a single unit CAM test is administered to each student. Each unit test contains four test items related to each of the six objectives just completed. The first two digits of the three digit test form number contain the unit which the form is measuring. Thus Test Form 115 is given immediately following the completion of Unit 11 and contains four test items on each of the six objectives in Unit 11; Test Form 125 is given immediately following the completion of Unit 12; etc. The relationship of test administrations to test forms for the unit CAM model is displayed in Figure 2.1.

The unit CAM test forms (115-245) each contain objectives related to one and only one unit. The student who responds to each of these unit tests during the instructional program will be tested by four items on all 48 objectives during Test Administrations 2-9. However, he will be tested on each objective during only one test administration (i.e., on the "posttest" for the unit). The six objectives in Unit 11 are measured on a short-term postinstructional basis (i.e., at completion of the unit). Similarly, when Test Form 125 is given to a student, the six objectives in Unit 12 are measured on an immediate postinstructional basis. Thus, during Test Administrations 2-9, information concerning the student's postinstructional achievement levels is gathered. There is no preinstructional (i.e., testing before the student has worked on the units) or retention (i.e., testing several weeks after the student has completed the unit) information provided by these unit levels.

FIGURE 2.1

The Relationship of Test Administrations

to Test Forms for the Unit CAM Model

Test Administration	Unit Completed	Test Form
1	none	1,2
2	11	115
3	12	125
4	13	135
5	14	145
6	21	215
7	22	225
8	23	235
9	24	245
10	all	1,2

The Sliding Unit CAM Model

The unit CAM model employs only one test form per test administration to gather detailed postinstructional information about each student. The unit CAM model supports some Level I decisions, but provides very little information for Level II decision making. The sliding unit CAM model uses multiple test forms during each test administration to gather some preinstructional and retention in addition to postinstructional information about the students and curriculum. The multiple test forms allow a student to take a form of the test, review objectives that he failed to learn, and retake a different form of the same test. The construction of these multiple test forms involves the use of stratified random sampling of objectives and test items to guarantee that the data generated by the evaluation program will be systematically related to the curriculum structure. Thus the sliding unit CAM models support both Level I and Level II decisions.

FIGURE 2.2

The Relationship of Test Forms to Test Administrations

in the Sliding Unit CAM Model

Test Administration	Unit Completed	Test Forms
1	none	1, 2
2	11	111, 112
3	12	121, 122
4	13	131, 132
5	14	141, 142
6	21	211, 212
7	22	221, 222
8	23	231, 232
9	24	241, 242
10	all	1, 2

Figure 2.2 contains the relationship of test forms to test administrations in the sliding unit CAM model. Forms 1 and 2, given at the beginning and end of the program, sample all the objectives in the course and represent a pretest/posttest component of the evaluation model. Forms 1 and 2 are discussed in the standard CAM model. During each of Test Administrations 2-9, two test forms are used that mostly measure the unit just completed. Figure 2.3 contains the actual test item numbers assigned to each question position of the sliding unit CAM test forms administered during the 7th test administration for each student. A test scheduling procedure is used to have half of the students repond to each form of the test.

In Figure 2.3 notice that 3 of the items on a form are related to Unit 21 (the last unit the student completed), 18 of the items are related to Unit 22 (the unit just completed), and the remaining 3 items are related to Unit 23 (the next unit to be worked on).

At this point a distinction needs to be made between a test form and a test. A test form is defined as several test items arranged in order. A test can be thought of as the set (meaning one or more) of

FIGURE 2.3

Items Assigned to Each Question Position of the Sliding Unit CAM Test Forms Used After Unit 22

Question Position	Items Assigned to Each Form 221	222
1	210103	210202
2	210305	210404
3	210601	210506
4	220101	220102
5	220103	220104
6	220105	220106
7	220202	220201
8	220204	220203
9	220206	220205
10	220301	220302
11	220303	220304
12	220305	220306
13	220402	220401
14	220404	220403
15	220406	220405
16	220501	220502
17	220503	220504
18	220505	220506
19	220602	220601
20	220604	220603
21	220606	220605
22	230103	230206
23	230405	230304
24	230501	230602

test forms that are related to the same curriculum content. A formal definition of test is presented later in the report. The test displayed in Figure 2.3 contains two test forms (221 and 222) each containing 24 test items. The technique of assigning test items from Units 21, 22, and 23 to Test Forms 221 and 222 is called stratified random sampling. Details concerning this and other sampling techniques can be found in Gorth (8).

Thus far, only the Unit 22 test has been examined. The tests related to the other seven units are similarly constructed. The tests related to each unit contain 48 items (two forms with 24 items each). Six of the 48 items are used to measure the last unit completed, 36 items are used to measure the unit being completed, and six items are used to measure the unit to be attempted next. The actual construction of each of these forms is similar to the specifications of Forms 221 and 222 shown in Figure 2.3.

There are advantages to using more than one form of a test for both Level I and Level II decision making. As stated above, a student can respond to a different form of the test to measure the success of his additional work. If only one form of a sliding unit CAM test were used then only three objectives in each unit could be measured on a preinstruction and retention basis. The use of two forms enables all objectives to be pretested and checked for retention achievement levels. This increase in information is important for curriculum revision and item reqriting decision (i.e., Level II decision making).

The Standard CAM Model

The standard CAM model is an evaluation technique for generating good information for Level II decision making. Correspondingly, the information generated for Level I decision making is not as good as that generated by the unit and sliding unit CAM models. Extenstions of the standard CAM model are applicable for Level III decision making evaluation designs (8). The standard CAM model for the hypothetical course consists of ten comprehensive interchangeable forms containing 24 items each. These forms are comprehensive in the sense that each one uniformly covers objectives in all eight units, and are interchangeable in that they are ten different forms of a 24-item final examination for the course. The items on each of the test forms are presented in Figure 2.4. Note that the forms are numbered 1 to 10. Each of these forms contains three items related to each of the eight units, and every item on a form is related to a different objective. Stratified random sampling was used to first assign the objectives to the question positions. The stratification process guarantees that three items per form are related to each of the eight units. Item sampling was then used to select the actual test items to be assigned to each question position.

At the beginning of the course (Test Administration 1), each student responds to one of these ten forms, and each form is responded to by 24 students in the course. During the second test administration,

FIGURE 2.4

Items Assigned to Each Question Position of Each of the Standard CAM Test Forms

Question Position	Items Assigned to Each Form 1	2	3	4	5	6	7	8	9	10
1	110106	110201	110102	110205	110204	110103	110104	110203	110206	110105
2	110403	110306	110302	110406	110301	110405	110305	110401	110304	110402
3	110506	110605	110604	110501	110603	110504	110502	110602	110503	110601
4	120202	120101	120203	120105	120104	120204	120201	120103	120206	120102
5	120303	120406	120306	120402	120401	120305	120405	120301	120404	120302
6	120505	120605	120501	120604	120506	120603	120602	120503	120606	120504
7	130106	130202	130103	130201	130205	130104	130204	130105	130206	130101
8	130404	130306	130405	130303	130302	130401	130402	130301	130403	130305
9	130501	130601	130502	130606	130503	130604	130504	130603	130602	130505
10	140103	140203	140104	140206	140201	140105	140205	140106	140204	140102
11	140405	140305	140401	140304	140303	140402	140302	140406	140301	140404
12	140502	140602	140503	140601	140506	140606	140505	140604	140501	140603
13	210104	210206	210203	210106	210101	210202	210102	210201	210103	210205
14	210301	210401	210405	210302	210404	210303	210306	210406	210403	210304
15	210603	210503	210604	210502	210605	210501	210606	210505	210602	210504
16	220105	220205	220105	220204	220203	220102	220103	220206	220104	220201
17	220302	220402	220306	220401	220304	220405	220406	220305	220403	220301
18	220604	220504	220503	220605	220601	220502	220602	220501	220506	220603
19	230201	230106	230202	230104	230203	230103	230206	230102	230205	230101
20	230302	230402	230401	230303	230304	230405	230305	230404	230301	230406
21	230504	230604	230505	230603	230501	230602	230606	230506	230605	230503
22	240101	240201	240102	240205	240206	240208	240208	240204	240202	240105
23	240403	240303	240302	240404	240301	240405	240306	240401	240304	240402
24	240505	240605	240606	240604	240502	240603	240503	240606	240601	240504

the process is repeated, but each student respond to a different test form. At the end of the instructional program, each student has responded once and only once to each of the ten test forms. Details concerning this scheduling process can be found in Gorth (8) and Pinsky (9).

Consider a student responding to a standard CAM test form following completion of Unit 22 (i.e., Test Administration 7). There are only 3 test items related to Unit 22 on each of these forms. Therefore, the standard CAM test does not provide information for deciding if the student has learned Unit 22 and should move on to Unit 23. Level I decisions are not supported by the standard CAM model. On the other hand, when the data from this standard CAM modal is summed over all the students in the course, it provides excellent data for Level II decision making. One can examine each objective in the course for input (preinstruction) and output (retention) achievement levels. The interaction of learning one objective upon the achievement levels of other objectives can be examined (for more detail see Gorth (9)).

Generalized Evaluation Concepts

Set - One or more. For example, a set of objectives is one or more objectives; a set of forms is one or more forms; a set of tests is one or more tests.

Content Span - A collection of ordered objectives specified in terms of the first and last objectives in the collection. In most instances the ordering of objectives is defined by the order in which they are taught. The name given the collection is related to the portion of the curriculum covered by the objectives in terms of content in text or time. For example, in the hypothetical curriculum, the content span contained Objectives 1101-1106 is Unit 11; the content span contained Objectives 1101-2406 is the entire curriculum.

Test Form - A collection of items in an order that is presented to the students. The term "form" is an accepted short version of the term "test form". Since each item is associated with an objective, the item numbers represent a content span for the test form.

Test - A set of forms that contains all test forms with the same content span.

Objective Density - The proportion of items related to an objective on a test or on a test form. The denominator of the objective density is the total number of items on the test or test form; the numerator is the number of items related to the specific objective on the test or test form.

Evaluation Period - A set of test administrations.

Standard Evaluation Component - A test consisting of more than one form that is used for more than one test administration during the evaluation period.

Sliding Unit Evaluation Component - A set of tests such that the content span of each test contains one or more objectives from the content span of the test used in the immediately preceeding or the immediately following test administration. Each test is used only once in an evaluation period.

Evaluation Design - A set of all the evaluation components designed for an evaluation period.

The unit CAM model is an evaluation design with two components - a standard component during Test Administrations 1 and 10, and a unit component during Test Administrations 2-9. The standard component consists of Forms 1 and 2 in Figure 2.4. This standard CAM test is based upon the content span of the entire curriculum, i.e., Objectives 1101-2406. Each of these objectives appears once on the test, and the objective density in the test is 1/48 for each of the 48 objectives. The unit component of the unit model consists of eight tests, each related to one of the eight units in the curriculum. The content span of the Unit 22 test (i.e., Form 235) is Objectives 2201-2206, the content span of the Unit 23 test (i.e., Form 235) is Objectives 2301-2306, and the content span of the Unit 24 test (i.e., Form 245) is Objectives 2401-2406. Note that the content span of these tests do not overlap, i.e., contain the same objectives.

The sliding unit CAM model is an evaluation design consisting of two components--a standard component in Test Administrations 1 and 10, and a sliding unit component in Test Administrations 2-9. The standard component of the sliding unit model is identical to the standard component of the unit model.

The sliding unit component consists of eight tests, one test related to each of the eight units. The test for Unit 22 has a content span of Objectives 2101-2306 (refer to Figure 2.3). This test contains 48 items, 24 items per form. Objectives 2101-2106 each appear only once on the test and have an objective density of 1/48; Objectives 2201-2206 each appear six times on the test and have an objective density of 6/48; Objectives 2301-2306 each appear once on the test and have an objective density of 1/48. The test for Unit 21 (not presented in the report) has a content span of Objectives 1401-2206. Objectives 1401-1406 each appear only once on the test and have an objective density of 1/48; Objectives 2101-2106 each appear six times on the test and have an objective density of 6/48; Objectives 2201-2206 each appear once on the test and have an objective density of 1/48 on the test.

The standard CAM model is an evaluation design consisting of a single standard evaluation component. The standard test (refer to Figure 2.4) consists of ten test forms defined over the entire curriculum content span (Objectives 1101 through 2406). The content span contains 48 objectives, and each objective appears five times on the test. Thus the objective density is 5/240 for each objective in the curriculum.

A Comparison of the Three Evaluation Models

Level I Decisions - The unit CAM model provides the most information for making decisions concerning an individual student's mastery of objectives on an immediate postinstruction basis. After each unit is completed the unit model generates four responses to each of the last six objectives completed. However, the unit CAM model does not provide any information concerning the student's preinstruction or retention achievement levels on the objectives in the course.

FIGURE 2.5

Analysis of Responses to Questions Measuring
Achievement on Objective 2201 by Test Administration

Model	Statistic	Test Administration 1	2	3	4	5	6	7	8	9	10
	Number of Responses	120	120	120	120	120	120	120	120	120	120
Standard	Percentage of Total Responses	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%
	Number of Items Used	5	5	5	5	5	5	5	5	5	5
	Number of Responses	120	0	0	0	0	120	720	120	0	120
Sliding Unit	Percentage of Total Responses	10%	0%	0%	0%	0%	10%	60%	10%	0%	10%
	Number of Items Used	1	0	0	0	0	1	6	1	0	1
	Number of Responses	120	0	0	0	0	0	960	0	0	120
Unit	Percentage of Total Responses	10%	0%	0%	0%	0%	0%	80%	0%	0%	10%
	Number of Items Used	1	0	0	0	0	0	8	0	0	1

The sliding unit CAM model provides very good information for mastery decision making relative to the individual student. This model generates three responses to each of the last six objectives completed. In addition, the sliding unit model provides a sampling of preinstruction and retention achievement levels for the individual student. In comparison with the unit model, the sliding unit model sacrifices some reliability concerning the immediate postinstruction mastery decisions (three items per objective rather than four) in order to gain some information concerning the student's preinstruction and retention achievement levels.

The standard CAM model provides very little information concerning the student's mastery of objectives. During each test administration, the standard model generates at most one response to the objectives in the course for each student. However, this model does provide an estimate of the student's preinstruction and retention achievement levels across all objectives every two weeks.

Level II Decisions - The unit CAM model provides very little information for Level II decision making. During Test Administrations 2-9, the model only provides group information about the latest six objectives completed. There is no information concerning the student's preinstruction and retention achievement levels on the other objectives in the curriculum.

The sliding unit CAM model provides more information about groups of students than the unit model. In addition to the group achievement level on the latest six objectives completed, the model generates estimates of group achievement on the previously completed six objectives and on the six objectives to be studied next. No information is provided concerning objectives completed more than one unit previously, nor concerning objectives to be studied after the next unit.

The standard CAM model provides information concerning the student's achievement on all 48 objectives following each test administration. Thus one is able to measure preinstruction, postinstruction, and retention achievement on all objectives in the course, and is able to measure the interaction effect of studying one objective upon the achievement levels of other objectives.

The information for groups of students generated by the three models is summarized in Figures 2.5 and 2.6. Figure 2.5 contains an analysis of the information about Objective 2201 generated by the three models in each test administration. Number of Responses refers to the number of possible student responses in each test administration to items related to Objective 2201. The hypothetical curriculum structure and evaluation models were designed such that each objective is responded to 1200 times by students during the program. It is the distribution of these 1200 responses over the test administrations that differs from model to model. This distribution is given by the row Percentage of Total Responses. The Number of Items Used refers to the degree of item sampling that is used in each model. For instance, in Test Administration 1, all models pro-

FIGURE 2.6

Analysis of Responses to Questions Measuring Achievement on Objective 2201 by Time Reference

Model	Statistic	Preinstruction	Postinstruction	Retention
Standard	Number of Responses	720	240	240
	Percentage of Total Responses	60%	20%	20%
Sliding Unit	Number of Responses	240	840	120
	Percentage of Total Responses	20%	70%	10%
Unit	Number of Responses	120	960	120
	Percentage of Total Responses	10%	80%	10%

Note: PREINSTRUCTION = Student response before instruction on Objective 2201

POSTINSTRUCTION = Student response to Objective 2201 during Test Administration 7 and 8.

RETENTION = Student response to Objective 2201 during Test Administration 9 and 10.

duce 120 student responses to Objective 2201. However, the standard model uses five items (ten forms are used), while the sliding unit and unit models use only one item each (only two forms are used). Remember that an objective only appears on every other form in the standard evaluation component. This figure displays the fact that the unit and sliding unit models generate more postinstruction information, while the standard model generates more preinstruction and retention information. Remember that each student completes instruction on Objective 2201 prior to Test Administration 7.

Figure 2.6 contains an analysis by time reference of the information about Objective 2201 generated by the three models. The student responses are broken down into PREINSTRUCTION (a response before instruction on the objective), POSTINSTRUCTION (a response to the objective during Test Administrations 7 and 8), and RETENTION (a response to the objective during Test Administrations 9 and 10). Note that as in Figure 2.5, each model generates 1200 student responses throughout the program. It is the distribution of responses over the time references (Percentage of Total Responses) that changes from model to model.

The power of the standard CAM model for Level II decision making can be seen if one considers the question as to the proper sequencing of objectives in the curriculum structure. By providing estimates of the class's achievement level ten times during the course (i.e., a longitudinal achievement measure), the standard model enables the teacher to recognize interactive instructional effects. Suppose that the instructional activities related to Objective 1203 also affect the achievement level on Objective 2201. Instruction on Objective 1203 during Test Administrations 2 and 3, but also provides an estimate of the students' achievement on Objective 2201 during these test administrations (see Figure 2.5). If there is a significant change in the achievement level on Objective 2201 from Test Administration 2 to Test Administration 3, the course structure might be resequenced the following year to include Objectives 1203 and 2201 in the same unit. An analysis of Figure 2.5 shows that the sliding unit and unit models generate virtually no longitudinal data.

Consider an input-output analysis of the effectiveness of the hypothetical course structure. Input is taken to mean the students' prein struction achievement level, and output is taken to mean the students' retention achievement level. Retention is being used as the output measure because postinstruction achievement levels sometimes contain transient achievement such as rote memory. An analysis of Figure 2.6 indicates that the standard model generates 80% of the student responses on a preinstruction and retention basis, the sliding unit model 30%, and the unit model 20% on a preinstruction and retention basis. Thus, the standard model generates data that is more useful for an input-output analysis of a course.

3. The Application of Computer-Assisted Instruction to Criterion-Referenced Testing

The criterion-referenced testing (CRT) evaluation designs discussed in the previous section can be classified as static testing. The tests are constructed before the students enter the program. The students are measured by these tests that cannot account for individual differences. Dynamic testing, on the other hand, would be able to construct each test based upon a student's past history. This type of testing would undoubtedly involve the use of a computer to print out individualized tests based upon the student's performance history that is stored in the computer. Computer-assisted instruction (CAI) uses dynamic testing in that it prints out test items or exercises based upon the student's level in the program. Of course, CAT does this testing on-line, and by providing immediate feedback on each item serves as an instructional as well as a testing mechanism.

A common structure for CAI is the strands structure that is used at Stanford University. For example, the elementary mathematics curriculum structure developed by Patrick Suppes contains 15 strands. Each strand includes all problem types of a given concept (e.g., fractions, equations) or of a major subtype of a concept (e.g. horizontal addition, vertical multiplication) presented in grades one through six. Within each strand, problems of a homogeneous type (e.g., all horizontal addition problems with a sum from zero to five) are grouped into equivalence classes. Each strand contains either five or ten classes per half-year with each class labeled in terms of a grade-placement equivalent.

A student is working on one equivalence class in each strand. The equivalence classes are structured in an increasing order of difficulty within each strand. Thus the student works on a given class until he passes a criteria after which he moves up to the next class. There are review exercises within a strand that the student must successfully respond to. Failure to correctly answer these review exercises can result in his being lowered a few equivalence classes. During each session at the computer terminal, the student responds to exercises from several strands. The emphasis placed on each strand depends upon the student's approximate grade placement, and upon his distribution of equivalence classes across the strands. A student will tend to receive more items on the strands where he is in the lower equivalence classes.

The major drawback to CAI has been the cost of having the students responding on-line to a computer. This report explores the possibility of using the CAI curriculum structure of strands and equivalences, printing the tests on the computer, and having the students respond to the tests off-line, i.e., at their desk in the classroom. There are many problems to this concept, and the report discusses these problems that need to be explored with more field-oriented research.

The equivalence classes within the strands structure are analogous to the performance objectives that are required in CRT. Many levels of

ERIC

performance objectives have been defined in the field of CRT. The equivalence classes most approximate enabling objectives are defined by O'Reilly (10). The generation of the test items can become a serious problem. In the mathematics CAI strands program, the test items are produced by item generation algorithms. Thus the test items themselves are not stored in the computer. Attempts to develop item generation algorithms in other subject matter areas have not been very successful. One of the major research areas in CRT today is producing useful item generation rules in a variety of subject matter areas (11).

An alternative to item generation rules is to physically store thousands of test items in disc storage and actually retrieve the items when required. Robert O'Reilly of the New York State Education Dept. has tried this technique for reading grades 4-6 and encountered serious problems. First, the computer software development costs were very high. The software includes adequate editing capabilities for correcting and modifying the item data base during the first year of operation. The data enty costs were extremely expensive. O'Reilly wanted to use upper-lower case characters and decided to enter his data via an optical character reading (OCR) machine. Problems were encountered when attempting to maintain quality control on the test item data base. Additional problems arose when test items required special symbols or diagrams. Computerized microfilm has the potential for solving many of the problems associated with maintaing a computerized test item bank. However, using today's technology, the microfilm technique is too slow and expensive.

Another problem that arises in attempting to apply CAI techniques to CRT is modifying the decision rules for moving a student through the equivalence classes or the analogous objectives. A student in the CAI program responds to approximately 50 test items every day. A student in a CRT program might respond to 50 test items per week. Thus there is only 20% as much information in the CRT program as in the CAI program. The ability to make decisions regarding changes in equivalence classes in each of the 15 strands following every test administration for a student becomes questionable. One runs into a classical bandwidth-fidlity measurement dilemma (12). As the amount of data decreases, the reliability of the decisions that are made decreases.

In the above paragraph, it is stated that a CRT program might generate responses to 50 test items per week (say one 50 item test per week). The reader might wonder why the CRT program could not include a test every day for each student. The problem lies in the cost factor. The two most expensive aspects of data processing today are input and output. The cost of producing a 50-item test for each student once a day is expensive. And the cost of entering student responses on a daily basis can become prohibitive. In addition, there is the cost (sometimes hidden) of administering these paper and pencil tests. The cost of on-line computer power is decreasing very rapidly these day, much faster than the cost of input-output devices. The cost of human clerical help is increasing year after year. Therefore, any attempts to monitor student progress on a daily basis might best be done using an on-line testing environment.

4. Earth Science Criterion-Referenced Testing Results

Background

During the fall of 1972 an experiment was conducted to determine the effectiveness of various uses of criterion-referenced testing and the ability of students to make their own instructional decisions based upon the CRT computerized output. The course used in the experiment was a 9th grade earth science course at San Carlos High School in San Carlos, California. The course was under the leadership of Larry Wagner at San Carlos High. The experiment was conducted with the cooperation of John Easter, Director of Project CAM, Sequoia Union High School District, Redwood City, California (13). There were originally four classes totalling 120 students involving two teachers in the study. However, one teacher left the school during the year. Thus the data presented here represents three classes, eighty-five students and one teacher. All results presented are based upon a small sample size and must be considered as tentative.

Each student at San Carlos High must complete one year of a science course before graduating. In the school, the college-bound students tend to take life science courses. The earth science program attracks a wide variety of ability and motivational level students. Data is presented concerning the background of the students in the program.

During the summers of 1969 and 1970, Wagner developed individualized study packets containing performance objectives, learning activities, self-tests, and posttests. During the 1970-71 school year students were free to select packets within each of the earth science content areas of astronomy, geology, meteorology, and oceanology. Each content area lasted one quarter. Based upon student feedback and CAM data, Wagner decided to modify the course design and change the evaluation procedures for the school year 1971-72.

The course design for 1971-72 was more traditional in that the students were group-paced but still used the packets developed for 1969-70. CAM results indicated a greater increase in student performance than in the previous year. There were still students who wanted to move independently and they were given the option. However, the number of students working independently was kept to a minimum to facilitate record keeping.

Keeping in mind that the group paced instruction data showed greater increases in student performance than the self-paced instruction group of the previous year, it was recognized that many students need to assume more responsibility and make more of their own decisions as to lesson selection and domplotion. Therefore, the format for the 72-73 school year included a group/self-paced combination as described below.

Curriculum Structure

The astronomy section of the earth science course is composed of 88 instructional objectives which make up 23 lessons within the six astronomy

SAN CARLOS EARTH SCIENCE

NAME ____________________

DATE ______________ PERIOD ______

Chapter 26 - STARS AND GALAXIES

LESSON 6 - STELLAR EVOLUTION AND GALAXIES

Objective Number INSTRUCTIONAL OBJECTIVES

2661 - Identify the correct description and/or size of our galaxy.

2662 - Associate the name of the 3 main types of galaxies to a description or diagram.

2663 - Identify the density, composition and origin of the great gas and dust clouds of interstellar space.

2664 - Be able to identify characteristics of each stage in the life history of a star in terms of temperature, color, and size of the star, and relate the process to our sun.

2665 - Be able to select the correct explanations or sketches which stand for the following origins of the universe:
a. Expanding universe theory
b. Steady-state theory

ACTIVITY OBJECTIVES

2666 (5) - Complete the study guide on STELLAR EVOLUTION AND GALAXIES using your text and other books as references.

2667 (5) - LABORATORY ACTIVITY - COMPARING THE SUN WITH OTHER STARS. Perform the activity as described in the handout with the same name.

2668 (5) - LABORATORY ACTIVITY - INVESTIGATING GALAXIES. Perform the activity as described in the handout with the same name.

2669 (20) - ESCP READINGS FOR GREATER UNDERSTANDING

a. STELLAR EVOLUTION (10) - Read pages 536-543 and answer questions 1-5 page 544.

b. WE LIVE in a GALAXY (5) - Read pages 544-547 and answer questions 1-4 page 547.

c. OUR GALAXY AMONG ITS NEIGHBORS (5) - Read pages 547-549 and answer questions 1-4 page 549.

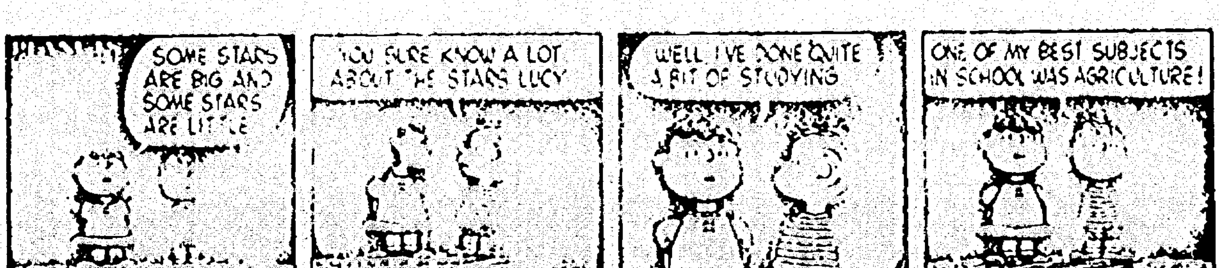

Figure 4.1 Example of objectives and activities for Lesson 6, Chapter 26.

chapters of the text. There are 2-7 lessons per chapter with an average of 3.8 objectives per lesson (from 2-5 objectives per lesson). Examples of the objectives are shown in Figure 4.1. The lessons are also made up of several activity objectives as shown in Figure 4.1.

The students were divided into three groups for the study, each group corresponding to a class section. The course was group-paced in terms of what chapters the students were studying, but was individually paced in terms of lessons (order and number) within the chapter. The major differences between the three groups were:

Group 1 - The teacher, based upon results of the CAM and Dubins Earth Science test, decided which lessons the student would study, and the lesson or lessons he had completed.

Group 2 - The student decided which lessons he would study and when he had completed a lesson.

Group 3 - The teacher, based upon results of the CAM and Dubins Earth Science test, decided which lessons the student would study, but the student decided when he had completed a lesson.

In all instances where a student declared that he had completed a lesson, he received positive credit for completing objectives within a lesson on which he answered the test items correctly. If he got these items wrong, he lost credit. The loss of credit policy was implemented to reduce the number of students who would have declared a lesson completed hoping that they would have gotten items correct by chance and thus recieved credit for objectives completed. In addition, a student was allowed to repeat any test (a different form). However, the score on the second test replaced the score on the first test even if the second score was lower. Thus, the student hopefully was motivated to study additional material if he chose to retake a test.

Evaluation Design

The evaluation design for the study consisted of CAM testing; two administrations of the Dubins Earth Science Test, one at the beginning and one at the end of the course; and two administrations of a student questionnaire, again once at the beginning and once at the end of the course. In addition, the results of the CTBS reading and mathematics tests for the students were obtained from the district records.

The CAM evaluation design consisted of a standard CAM with two forms given on Test Administrations 1 and 8, and sliding unit CAMs given during Test Administrations 2-7. The two forms of the standard CAM were made up of questions randomly selected from the astronomy bank of items so that all lessons were sampled. Each student received the form during Test Administration 8 that he did not receive during Test Administration 1. The sliding unit CAM contained two items per lesson on lessons already completed; five items per lesson on the lessons just studied; and two items per lesson on the lessons to be studied next. The sliding unit CAMs were administered every two to three weeks. Each test consisted of two forms so that one-half of the class received each form. If a student chose to retake a test, he took the alternative form.

The Dubins Earth Science Test consists of two forms (A and B) each containing 60 items. Form A was given in September, Form B given in January. The test is divided into four content areas, Geology, Astronomy, Meteorology, and Oceanography; and three content distributions, knowledge, understanding, and application. There are 31 items related to Astronomy; and 47 knowledge items, 33 understanding items, and 50 application items on the test.

The student questionnaire consisted of two 30-item forms. Both forms were used at the early October and mid-December administrations of the questionnaire. The students who took Form 1 during October, responded to Form 2 during December and vice-versa. The questionnaire consisted of statements that the student was asked to agree or disagree with on a five point scale. Several items were worded negatively to increase the validity of the instrument. All data were processed so that a response of 1 indicated the most positive agreement and a response of 5 indicated the most positive disagreement. All items on the questionnaire were divided into six categories. These were (1) attitude on content and activities; (2) attitude on decision making; (3) test anxiety; (4) course anxiety; (5) self-concept; and (6) use of CAM data. The questionnaire items were constructed and catagorized by members of the High School District.

Operating Procedures

A major problem encountered in the study was how to convert the student decisions concerning lesson completion to a computer readable format. The computer output shown in Figure 4.2 was designed to overcome this problem. Before the student responded to a test form, he was given a computerized lesson completion summary sheet. This sheet contains the date of lesson completion (a blank means that he has not completed the lesson), his preinstruction and postinstruction scores on the lesson. He circled those lessons that he had recently completed. The sheet was sent to keypunching for input into the computer.

While the above technique worked well for the 85 students in the study, it would prove quite expensive on a large scale basis. Other techniques of entering lesson completion data need to be developed for individually paced instructional programs.

Results

Figures 4.3 and 4.4 contain summarized results of the data collected in the study. All data (except the CTBS scores) were generated by the CAM2 computer software run on a Hewlett-Packard 2120 computer system at the Sequoia Union High School District Central Office. Figure 4.3

FIGURE 4.2

Lesson Completion Summary

KRAMER SCOTT L 105981 WAGNER PERIOD 2

11/29/72 SC202

LESSON	DATE COMPLETED	PRE-INS RESPONSES	POST-INS RESPONSES
261	9/22/72	0/ 2	7/ 7
262	9/22/72	2/ 2	7/ 7
263	9/22/72	0/ 2	6/ 7
264		6/ 9	0/ 0
265	10/ 6/72	1/ 4	4/ 7
266	10/ 6/72	2/ 5	6/ 7
267		7/11	0/ 0
271	10/25/72	1/ 5	7/ 7
272	10/25/72	4/ 4	7/ 7
273	10/25/72	2/ 4	6/ 7
274	10/25/72	1/ 4	7/ 7
275	10/25/72	1/ 4	6/ 7
291	11/ 9/72	2/ 4	4/ 7
292	11/ 9/72	3/ 4	7/ 7
293	11/ 9/72	0/ 4	7/ 7
294	11/ 9/72	0/ 4	6/ 7
301	11/29/72	1/ 4	5/ 5
302	11/29/72	4/ 4	5/ 5
303	11/29/72	2/ 4	4/ 5
311		1/ 5	0/ 0
312		2/ 4	0/ 0
321		2/ 4	0/ 0
322		2/ 5	0/ 0
TOTALS		46/102	101/113

TOTAL POINTS 178

contains the results displayed for all students in the course and for each of the three periods, while Figure 4.4 contains the data displayed by student grade levels, i.e., those students who received an A in the course, those that received a B, and those that received a C.

Figure 4.3 contains the CTBS reading and mathematics scores of the students in each of the three periods. These scores are in terms of the national percentiles for 9th grade students during the month of October. The results indicate that Periods 1 and 2 entered the course with approximately the same achievement background in reading and mathematics, while Period 3 was significantly lower in both these areas. The average course grade for Period 1 was 3.0 (3.0 = B), Period 2 was 2.6 and Period 3 was 2.4 (2.0 = C). Moreover, the average student in Period 1 recieved 4.1 units of credit, the average in Period 2 was 4.0, while the average in Period 3 was 3.8. Full credit for the semester's work was 5.0.

Based upon the standard CAM test given at the beginning and end of the course, Period 2 had the highest entry level (32% correct) and the lowest gain in achievement. Period 1 has the highest gain (57-29=28%) Period 3 the second highest (51-25=26%), followed by Period 2 (55-32=23%). Period 2 also had the smallest gain on the astronomy portion of the Dubins Earth Science test (55-41=21%) and Period 3 (53-25=58%). It is interesting to note that Period 3 made the gain on the Dubins primarily in the understanding and application components of the test.

The student questionnaire data shows little differences between the three periods. Periods 2 and 3 become less positive toward the course at the end of the semester (2.5 and 2.6 and 2.5 and 2.8). Period 3 students responded slightly negatively (3.2) to test anxiety and use of CAM data questions during the second administration of the questionnaire.

Figure 4.4 contains the results displayed by course grade level. As would be expected, the CAM results are much higher for the A students (71% postinstruction) than for the C students (45% postinstruction). However, the gain in achievement on the astronomy part of the Dubins test were approximately the same for all three grade levels. This gain seems to have been made mostly on the understanding and application components of Dubin, especially for the C students.

The questionnaire results tend to show that the A and B students thought more favorably of the course at the end than did the C students. This difference appears to be spread over all categories of questions.

Discussion of the Results

There are several problems with interpreting the results of this study. The first is the sample size. Much of the data presented in Figures 4.3 and 4.4 are based upon small enough sample sizes to create doubt about any statistical significance between pairs of values. One must rather look at trends in the data over several measures. Secondly, the study was conducted in a real life high school environment. The

FIGURE 4.3

Results Displayed by Course Periods

Instrument	All Students		Period 1		Period 2		Period 3	
CTBS Reading[1]	53		57		56		46	
CTBS Mathematics[1]	43		50		48		32	
Course Grade[2]	2.7		3.0		2.6		2.4	
Units of Credit	4.0		4.1		4.0		3.8	
Number of Students	85		30		30		25	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Standard CAM[3]	29	54	29	57	32	55	25	51
Dubins Total[3]	35	43	37	45	36	41	30	43
Dubins Astronomy[3]	36	56	38	59	41	55	25	53
Dubins Knowledge[3]	35	38	37	42	36	37	30	36
Dubins Understanding[3]	38	47	42	50	44	43	26	50
Dubins Application[3]	33	45	35	45	31	43	32	46
Attitude on Content[4]	2.6	2.6	2.5	2.5	2.6	2.6	2.6	2.7
Attitude on Decisions[4]	2.4	2.5	2.4	2.5	2.4	2.3	2.4	2.6
Test Anxiety[4]	3.0	3.0	3.1	2.9	2.9	3.0	2.9	3.2
Course Anxiety[4]	2.2	2.3	2.0	2.2	2.2	2.3	2.2	2.5
Self-concept[4]	2.2	2.5	2.3	2.4	2.2	2.4	2.2	2.5
Use of CAM Data[4]	2.7	3.0	2.7	2.7	2.8	3.0	2.6	3.2
Questionnaire Average[4]	2.5	2.7	2.5	2.5	2.5	2.6	2.5	2.8

Notes: 1: expressed in national percentiles
2: A=4.0, B=3.0, C=2.0, D=1.0
3: expressed as the percentage of correct responses
4: 1.0=strongly agree,.... 5.0=strongly disagree

FIGURE 4.4

Results Displayed by Course Grade Level

	Student Group					
	A Students		B Students		C Students	
Number of Students	18		24		42	
Instrument	Pre	Post	Pre	Post	Pre	Post
Standard CAM[1]	33	71	31	61	26	45
Dubins Total[1]	47	58	38	46	29	36
Dubins Astronomy[1]	51	72	39	61	29	48
Dubins Knowledge[1]	44	52	37	40	30	32
Dubins Understanding[1]	52	66	41	46	32	41
Dubins Application[1]	47	60	37	49	24	36
Attitude on Content[2]	2.5	2.5	2.4	2.6	2.8	2.8
Attitude on Decisions[2]	2.4	2.3	2.4	2.4	2.4	2.7
Test Anxiety[2]	3.1	2.7	3.2	3.0	2.9	3.2
Course Anxiety[2]	2.3	2.2	2.1	2.2	2.3	2.5
Self-concept[2]	2.0	2.3	2.4	2.3	2.2	2.7
Use of CAM Data[2]	2.3	2.8	2.9	2.9	2.8	3.2
Questionnaire Average[2]	2.3	2.5	2.6	2.6	2.6	2.9

Note: 1: expressed as the percentage of correct responses
2: 1.0=strongly agree, . . ., 5.0=strongly disagree

teacher was not always able to maintain the conditions of the study. Sometimes he was not able to direct the study of each student in Period 1 due to a lack of time; while in Period 2, he was forced to abandon the study conditions to work with some students who would not have passed the course if left to their own study decisions. And thirdly, the value of the student questionnaire as an evaluation instrument is debatable. Experience with other student questionnaires in the Sequoia High School District indicates that students tend to be overly agreeable toward the teacher on these instruments.

Despite these problems, the data seem to strongly indicate that these earth science students need teacher support when using CRT data generated by the computer. The teacher felt that the data overwhelmed the students who were not quantitatively oriented. However, when using the data with the assistance of a teacher or paraprofessional, the students seemed to increase their achievement levels (this statement being based upon other studies done in past years in the same course (13)).

The study suggests a need for additional applied research on the use and effectiveness of CRT in the classroom. One research effort would be to replicate the study using different subject matter areas and different student populations. Another research study would be to analyze the amount of training a student needs to benefit from computerized CRT data. Should computerized CRT reports be distributed to third grade students who have low quantitative abilities? Or should the results be presented to students by a teacher or paraprofessional? Is the student better able to utilize the CRT information during his second and third years in a CRT program?

REFERENCES

1. Committee for Economic Development, Innovations in Education: New Directions for the American School, New York, July, 1968.

2. Esbenson, Thorwald, Individualizing the Instructional Program, Duluth, Minnesota Public Schools, 1966.

3. Cooley, William, and Glaser, Robert, "An Information and Management System for Individually Prescribed Instruction," Working Paper 24, University of Pittsburgh, December, 1968.

4. Enterprise Elementary School District, "Enterprise Technologically Managed Individualized Instruction Program, "Enterprise, California, January, 1970.

5. Silberman, H.G., "Design Objectives of the Instructional Management System," SP 3038/001/00, Systems Development Corporation, Santa Monica, California, 1968.

6. Gorth, William, "Designing Instructional Systems with Longitudinal Testing Using Item Sampling Techniques," a symposium at the Annual Meeting of the American Educational Research Association in Minneapolis, March, 1970.

7. Lord, F.M., and Novick, M.R., Statistical Theories of Neutral Test Scores, Addison-Wesley, Boston, 1968.

8. Gorth, W.D., Schriber,P.E., and O'Reilly, R.P., Comprehensive Achievement Monitoring: Its Design and Use, School of Education, University of Massachusetts, Amherst, Massachusetts, 1971.

9. Pinsky, P.D., "Mathematical Models for Measurement and Control of Classroom Achievement," unpublished dissertation, The Department of Operations Research, Stanford University, Stanford, Cal., 1971.

10. O'Reilly, R.P., "The Conceptualization of Objectives for Evaluation," a paper presented at the Annual Meeting of the Educational Research Association in New Orleans, February, 1973.

11. Bormuth, J.R., On the Theory of Achievement Test Items, The University of Chicago Press, 1970.

12. Cronbach, L.J., and Gleser, G.C., Psychological Tests and Personnel Decisions, University of Illinois Press, Urbana, 1965, Chapter 8.

13. Easter, J.E., et. al., "Comprehensive Achievement Monitoring in the Sequoia Union High School District," a symposium presented at the California Educational Data Processing Association Meeting, Palo Alto, California, December, 1972.